Project no. 043338

Project acronym: EMERGENCE

Project title: A foundation for Synthetic Biology in Europe

Instrument: NEST Pathfinder

Thematic Priority: Synthetic Biology

Deliverable 1.4:

Report on recommendations of the intra-consortium expert group on suitable promoter standardization formats

Due date of deliverable: Dec. 2007
Actual submission date: May 2008

Start date of project: 1.12.2006 Duration: 36 months

Victor de Lorenzo
Centro Nacional de Biotecnologia, CSIC, Campus de Cantoblanco, 28049 Madrid, Spain

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Report on recommendations of the intra-consortium expert group on suitable promoter standardization formats[1]

Introduction. The basic idea behind Synthetic Biology is that any biological system can be regarded as a combination of individual functional elements—not unlike those found in man-made devices—and can thus be described as a limited number of parts that can be combined in novel configurations to modify existing properties or create new ones. In this context, engineering transits from being an analogy of the rational combination of genes—as in standard Molecular Biology and Biotechnology—to becoming a veritable methodology with which to construct complex biological systems from first principles. The fusion between authentic (not metaphoric) Engineering and Molecular Biology will certainly have far-reaching consequences, but to what extent is this realistic science; how much is genuinely new and how much is merely hype generated by rebranding existing things? According to long-standing philosophical tradition, Science is about knowing and understanding, while Technology is about doing (Wolpert, 1998). So, in what realm does synthetic biology fall? For many of its practitioners, the answer is clear: Synthetic Biology is about Engineering, not about science (Endy, 2005; Baker et al, 2006; Andrianantoandro et al, 2006). But engineers are not the only stakeholders as Synthetic Biology is attracting a large number of researchers from fundamental Science (Church, 2005) as well -let alone companies and businesses, although their agendas are quite diverse (Fig. 1).

The transatlantic debate on Biological parts. It is even possible to distinguish a unique European perspective, as many activities that now qualify as Synthetic Biology -protein design, modelling, metabolic engineering, biological nanomachines -have been going on for some time on the 'old continent'. In fact, many European scientists are rather skeptical about calling Synthetic Biology a brand new field as there is a clear similarity—though a different language—between the discourse of genetic engineering of the late 1970s and many of the claims and assertions made by synthetic biologists.

[1] Note that much of this report will be published as an opinion paper in EMBO Reports: **de Lorenzo, V. and Danchin, A.** (2008) Synthetic Biology: discovering new worlds and new words. *EMBO Rep* (In Press).

However, these various biological fields have always been more implicit than explicit, quite fractionated and lacking in a common descriptive language. In contrast, the present momentum for Synthetic Biology is a good opportunity to realize a common potential, find a shared language and identify synergies. In our view, the key to fulfilling the promise of Synthetic Biology—in terms of scientific *and* technological breakthroughs -is not about societal acceptance or ethics, but rather about understanding the biological building blocks that can be used for robust engineering; about adopting a descriptive, quantitative language for biological transactions; and identifying and managing the physical and chemical constraints that frame the functioning of any autonomous biological system.

Biological parts -minimal biological elements that can be used for engineering- are one of the trademarks of ongoing efforts in Synthetic Biology. The idea is both simple and attractive: in the same way that a machine can be disassembled and catalogued as individual components—such as hard disks, screens, keyboards or memory chips—living systems may also be broken down into a list of components that can be rewired for a specific purpose. This sounds like a straightforward engineering approach, but it may not be that easy. The functions of nearly all extant biological systems -living entities- depend on the environment in which they thrive and the evolutionary pressures that have created a growing complexity of interaction at all levels. Furthermore, proteins seem to possess an amazing ability to develop new interactions with other proteins as soon as they are subjected to selective pressure. We need a better conceptual frame to understand what minimal biological building blocks are and how they can be defined. Just calling them Biobricks™ and regarding them as singular biological components—as in the MIT-run catalogue of biological parts (http://partsregistry.org) -can give a misleading perception of the issues at stake. Furthermore, the nature and description of such parts depends on the scale of the engineering objective. While genetic circuits may rely only on defined promoters and reporters, designing a whole cell will require complete functional modules (translation, energy generation, replication etc) as building blocks. Similarly, whole cells will be the parts for microbial community design and tissue engineering, and so on.

Where should we get the building blocks from? The ultimate agenda of Synthetic Biology is to recreate a cell as an automaton that can algorithmically process information. To this end, we first need to identify the different functions of a cell before

compiling a list of the parts that implement such functions. An important point here is to avoid the significant trap of assuming any goal in such an automaton; all of its properties should be *declarative*, not *prescriptive*: there are no built-in instructions to tell the automaton what it should do. The comparative analysis of living organisms should give us a list of the functions that are needed for life. Such a research programme, however, might look hopeless from an engineering perspective as many different objects can fulfil the same function. Fortunately, evolution can help us to solve this problem: life evolves by ascending from earlier life forms and any function that has emerged and been implemented within or by a particular biological system becomes conserved over generations. This evolutionary 'stickiness' can be analyzed by identifying persistent genes, i.e. those that are recurrently kept in a given number of genomes (Fang et al, 2005). Using persistent genes -which are by no means expected to be ubiquitous- it is possible to construct a first catalogue of 400 to 500 functions that appear to be essential for life. Yet, a significant number of persistent genes have unknown functions, and we may miss other, essential ones. For instance, we may fail to identify functions associated with membranes, as the rules that define similarities between membrane proteins might be distinct from those of cytoplasmic proteins.

At least in the case of bacterial genomes, the global set of genes can be split into two categories: those that allow life and perpetuate it, and those that permit life in an environmental context. We call the class of persistent genes on the first list the *paleome*, the members of which constitute the list of minimal biological functions (Danchin et al, 2007). It is in such a list where we need to search for all components to be implemented in an artificial cell able to mimic the behaviour of living entities. The quest for a minimal set of functions for a self-maintaining system is not limited to Synthetic Biology. For some time, engineers have been working on a self-reproducible three-dimensional printer (http://www.reprap.org). Their work shows that a Turing Machine (Turing, 1937) could serve as a model for a synthetic living system that would contain the machine itself but also require a separate program to store a blueprint of how to assemble it. It would also need a source of energy and transport systems to capture missing parts from the environment and lubricants to permit the movement of components. The experiences gained from designing a self-reproducing printer provide a number of interesting lessons for the overall architecture of biological systems and the interactions between the parts. There is a take-home message: engineering biological

systems involves much more than cutting and pasting DNA sequences of more or less characterized parts -even if one can build on a logical blueprint.

One key issue: defining transcriptional units. Every descriptive language, including those used to describe technical or scientific systems, is ultimately metaphorical; it carries a meaning and has an agenda (Danchin, 2003). Although molecular biologists often believe that their abstractions and representations -many of them taken from Physics—are the ultimate means to represent biological phenomena, their language may not be sufficient to fulfil Synthetic Biology's strong engineering agenda. A robust language to describe engineering biological entities is seriously needed, but must be based also on sound Biology. Simply renaming longstanding concepts such as transcription or translation rates by equivalent terms to echo signal-transmission in electronic circuits may give a misleading perception of the issues at stake. For instance, a number of US Synthetic Biology groups (http://syntheticbiology.org) have adopted the term PoPS (polymerase per second) to quantify the input/output signals in genetic circuits. PoPS describes the flow of RNA polymerase molecules along DNA (i.e., the *current* for gene expression), so that PoPS level is set by the amount of molecules of the enzyme that go through a specific position on DNA each second. Similarly, RIPS (ribosome per second) refer to the flow of the translation machinery through mRNA. There is little Biology in these definitions, only a straight and overtly simplistic projection of electric engineering concepts into (supposedly) biological counterparts. Is this ultimately the way to go? This specific issue deserves some thought, as the challenge of describing and standardizing autonomous biological parts is not just academic. To achieve the engineering goals of synthetic biology, we need to adopt a consensus on robust engineer-able elements -like the ISO metric standards that are now universally accepted. In this context, we need to start with a quantitative standardization of the signal transduction between these parts, e.g. the transcriptional activity of distinct promoters *in vivo* and their quantification in universal units. But each scientist seems to have a favourite way of measuring such a value with all kinds of reporter genes or DNA chips, let alone a plethora of miscellaneous hosts, gene doses, media and temperatures, which must be replaced by unequivocal promoter strength units that engineers can use to calculate their circuits. This discussion must involve not only PoPS enthusiasts and synthetic biologists, but also experts in the fundamental aspects of transcription with all its intricacies. The definition of transcription units and many other types of biological

functions may eventually be subject to some governance in order to establish benchmarks. There are already discussions to promote a European Institute of Biological Standards as a counterpart of the MIT-run initiatives mentioned above. Yet, even if we have a set of standardized parts and functionalities, we may still lack the knowledge of how to rewire these -akin to writing a book with well defined words but lacking the grammar. One possible solution (the only one available so far) is to use biological chasses, extant or synthetic genomes, as sort of 'grey box' modules in which to implant characterized and predictable circuits.

Dealing with evolution. Biological entities are not only prone to become interdependent; they also evolve in unpredictable ways as they are subjected to the cycle of mutation/ amplification/ selection that is intrinsic to evolution. The implantation of extra DNA into a cell and the proteins encoded are severely counter-selected over time if they cause any burden to cell physiology. This notion is suggested by the long period of time that horizontally transferred genes take to develop regulatory interactions (Lercher & Pal, 2007) and by the problems encountered when transferring genes whose products belong to multiprotein complexes (Sorek et al, 2007). The practical downside of these biological phenomena is the difficulty to stably programme bacteria with genetic circuits or through heterologous expression of regulatory modules. Bacteriophages that had been redesigned to behave in a more logical way (Chan et al, 2005) made smaller lysis plaques than their wild type precursors and eventually evolve to erase the human construction parts. We therefore need to explore how to avoid or decrease undesired evolution. One possibility might be to use endogenous DNA repair systems to keep the fidelity of the instructions encoded in the implanted DNA. One can also think of engineering a minimum interference within the host by means of orthogonal parts. Ultimately, it is a question of whether an alternative information-coding molecule and the corresponding expression machinery can be produced to be less amenable to mutation than DNA. One could think about the other extreme and create highly evolvable biological modules with a capacity to nest rapidly in a pre-existing regulatory network (Silva-Rocha & de Lorenzo, 2008), reminiscent of the programs that install new software on the operating system of a computer.

Many synthetic biologists adopt the implicit or explicit metaphor of the cell as a complex mechanical machine, which requires relevant sub-machines to organize itself,

including scaffolds. How can we identify these components? A remarkable feature of the paleome (see above) is that these genes are systematically coded in the leading replication strand, which shows that there is strong selection pressure to avoid conflicts between transcription and replication (Fang et al, 2005; Rocha & Danchin, 2003). It is therefore important to compile a list of the corresponding objects, which, in engineering terms, would be sub-machines. A general way to identify these complexes is to analyze groups of co-evolving genes in the paleome, such as the genes that determine the buildup of the ribosome, for instance. Other examples would be the transcription "nanomachine" possibly coupled to the ribosome, the replication 'nanomachine' or the 'nanomachinery' that shapes the cell and organizes its division.

The third factor: metabolism. While genome provides a complete catalogue of genes, it is not yet possible to get a complete list of a cell's metabolites by analyzing its genome. However, metabolic transactions impose a chemical and energetic framework upon the cell, a sort of inescapable background economy. While the links between the transcription and translation of mRNA in the ribosome are well known, the organization of metabolism and its influence in controlling cell activities is much less so. Allosteric regulation of enzymes by intermediate metabolites, which was an important topic of biochemical research in the 1960s and 1970s, was virtually abandoned in favour of transcriptional regulation by protein factors and signal molecules. The question of how metabolites interface with the protein machinery that controls genetic networks is largely unexplored, but is certainly relevant for engineering biological circuits. But this should not deter us from pushing forward the Synthetic Biology agenda: such problems are perhaps not so different from the challenge of engineering an airplane, where hundreds of kilometres of cables, the circulation of kerosene, the maintenance of a correct atmosphere and temperature, control panels and devices, seats, lights etc. have all to be organized.

We thus advocate the metaphor of the cell as an algorithmic, rather than a mechanical machine, and the adequacy of machine-oriented engineering language to implement synthetic biology. Under this scheme, the roadmap to engineering biological systems is not determined by the biological parts, but by how they interact. As is the case for the 3D printer, the relationships between the objects—not necessarily the objects themselves—are absolutely central to any attempt to construct a synthetic cell endowed

with non-natural properties and utilities. This is implicitly accepted by each suggestion to replace a given biological part, such as using amino acids that differ from the 20 natural ones to construct proteins, for example.

Aging of biological systems. These observations point to a major challenge that has not been generally raised, although it has been discussed by engineers: Even if we construct a synthetic cell, its very functioning will make it age and wither (Nystrom, 2003; Nystrom, 2007). Again, a careful analysis of the paleome may help to solve this problem. Perusal of the most persistent genes shows that they are apparently dispensable for colony formation in the laboratory (Fang et al, 2005), most encode functions involved in maintenance and repair and are therefore involved in the perpetuation of life, rather than in permitting life *per se*. We think that this is an essential feature of living organisms that needs to be taken into account when constructing synthetic cells. Indeed, the prospect of making cells *à la carte* for industrial production calls for robust constructs that can easily be scaled up to large production volumes by cell divisions over many generations without altering the cells' properties and/or the decoupling of growth from catalytic performance. The separation of the paleome into two major functionalities is reminiscent of the necessary distinction between life perpetuation/construction and reproduction/replication (Dyson, 1985). While the latter enables life but accumulates errors, understanding the former can teach us to program long-lasting synthetic cells,

REFERENCES:

Andrianantoandro E, Basu S, Karig D, Weiss R (2006) Synthetic Biology: new engineering rules for an emerging discipline. Mol Sys Biol **2:**0028

Baker D, Church G, Collins J, Endy D, Jacobson J, Keasling J, Modrich P, Smolke C, Weiss R (2006) Engineering life: building a fab for biology. *Sci Am* **294**: 44-51

Chan LY, Kosuri S, Endy D (2005) Refactoring bacteriophage T7. *Mol Syst Biol* **1:** e0018.

Church GM (2005) From systems biology to synthetic biology. *Mol Syst Biol* **1:** e 0032.

Danchin A (2003) T*he Delphic Boat: What genomes tell us*. Cambridge, MA, USA: Harvard University Press,.

Danchin A (2008) A phylogenetic view of bacterial ribonucleases. *Progr Nucl Acids Res Mol* Biol (in Press).

Danchin A, Fang G, Noria S (2007) The extant core bacterial proteome is an archive of the origin of life. *Proteomics* **7:** 875-889.

de Lorenzo V, Perez-Martin J (1996) Regulatory noise in prokaryotic promoters: how bacteria learn to respond to novel environmental signals. *Mol Microbiol* **19:** 1177-1184.

Dyson FJ (1985) *Origins of life*. Cambridge, UK: Cambridge University Press

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* **403:** 335-338.

Endy D (2005) Foundations for engineering biology. *Nature* **438:** 449-453.

Fang G, Rocha E, Danchin A (2005) How essential are nonessential genes? *Mol Biol Evol* **22:** 2147-2156.

Feist AM, Palsson BO (2008) The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. *Nat Biotechnol* **26:** 659-667.

Fredriksson A, Nystrom T (2006) Conditional and replicative senescence in *Escherichia coli. Curr Opin Microbiol* **9:** 612-618.

Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. *Nat Rev Microbiol* **3:** 722-732.

Gibson DG et al (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science* **319:** 1215-1220.

Jansson JK (1995) Tracking genetically engineered microorganisms in nature. *Curr Opin Biotechnol* **6:** 275-283.

Lartigue C, Glass JI, Alperovich N, Pieper R, Parmar PP, Hutchison CA, 3rd, Smith HO, Venter JC (2007) Genome transplantation in bacteria: changing one species to another. *Science* **317:** 632-638.

Leduc S (1912) *La Biologie Synthétique*. Paris, France: Poinat

Lercher MJ, Pal C (2007) Integration of horizontally transferred genes into Regulatory interaction networks takes many million years. *Mol Biol Evol* **25:** 559-567.

Luisi PL (2006) *The emergence of Life. From chemical origins to Synthetic Biology*. Cambridge, UK: Cambridge University Press

Mazel D (2006) Integrons: agents of bacterial evolution. *Nat Rev Microbiol* **4:** 608-620.

Noireaux V, Libchaber A (2004) A vesicle bioreactor as a step toward an artificial cell assembly. *Proc Natl Acad Sci USA* **101:** 17669-17674.

Nystrom T (2003) Conditional senescence in bacteria: death of the immortals. *Mol Microbiol* **48:** 17-23.

Nystrom T (2007) A bacterial kind of aging. *PLoS Genet* **3**: e224.

Pedraza JM, van Oudenaarden A (2005) Noise propagation in gene networks. *Science* **307:** 1965-1969.

Qian H (2006) Reducing intrinsic biochemical noise in cells and its thermodynamic limit. *J Mol Biol* **362:** 387-392.

Ramos JL, Diaz E, Dowling D, de Lorenzo V, Molin S, O'Gara F, Ramos C, Timmis KN (1994) The behavior of bacteria designed for biodegradation. *Biotechnology* (NY) **12:** 1349-1356.

Rasmussen S, Chen L, Nilsson M,Abe S (2003) Bridging nonliving and living matter. *Artif Life* **9:** 269-316.

Rocha EP, Danchin A (2003) Gene essentiality determines chromosome organisation in bacteria. *Nucleic Acids Res* **31:** 6570-6577.

Shimizu T, Matsuoka Y, Shirasawa T (2005) Biological significance of isoaspartate and its repair system. *Biol Pharm Bull* **28:** 1590-1596.

Silva-Rocha R, de Lorenzo V (2008) Mining logic gates in prokaryotic transcriptional regulation networks. *FEBS Lett* **582:** 1237-1244.

Sorek R, Zhu Y, Creevey CJ, Francino MP, Bork P, Rubin EM (2007) Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318:** 1449-1452.

Tamames J, Gonzalez-Moreno M, Mingorance J, Valencia A, Vicente M (2001) Bringing gene order into bacterial shape. *Trends Genet* **17:** 124-126.

Turing AM (1937) On Computable Numbers, With an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* Series 2, Volume 42.

Wang K, Neumann H, Peak-Chew SY, Chin JW (2007) Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat Biotechnol* **25:** 770-777.

Wolpert L (1998) *The unnatural nature of science: why science does not make (common) sense*. Cambridge, MA, USA: Harvard University Press